



An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer–composite materials

Dinh Dong Doan, Emmanuel Ramasso, Vincent Placet, S. Zhang, Lamine Boubakar, Noureddine Zerhouni

► To cite this version:

Dinh Dong Doan, Emmanuel Ramasso, Vincent Placet, S. Zhang, Lamine Boubakar, et al.. An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer–composite materials. Mechanical Systems and Signal Processing, 2015. hal-01303380

HAL Id: hal-01303380

<https://hal.science/hal-01303380>

Submitted on 18 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 An unsupervised pattern recognition approach for AE
2 data originating from fatigue tests on
3 polymer-composite materials

4 D.D. Doan^a, E. Ramasso^{a,b,*}, V. Placet^b, S. Zhang^b, L. Boubakar^b, N.
5 Zerhouni^a

6 ^a*Department of Automation Science and Micro-Mechatronic Systems*

7 ^b*Department of Applied Mechanics*

8 *FEMTO-ST, 32 avenue de l'Observatoire, 25000 Besançon, France*

9 **Abstract**

This work investigates acoustic emission generated during tension fatigue tests carried out on a carbon fiber reinforced polymer (CFRP) composite specimen. Since fatigue data processing, especially noise reduction remains an important challenge in AE data analysis, a Mahalanobis distance-based noise modeling has been proposed in the present work to tackle this problem. A Davies-Bouldin-index-based sequential feature selection has been implemented for fast dimensionality reduction. A classifier offline-learned from quasi-static data is then used to classify the processed data to different AE sources with the possibility to dynamically accommodate with unseen ones. With an efficient proposed noise removal and automatic separation of AE events, this pattern discovery procedure provides an insight into fatigue damage development in composites in presence of millions of AE events.

10 *Keywords:* organic-matrix composites, acoustic emission clustering, fatigue
11 datasets, noise reduction, sequential feature selection.

*Corresponding author. Tel.: +33 81 66 69 49.

Email address: emmanuel.ramasso@univ-fcomte.fr (E. Ramasso)
Preprint submitted to Mechanical Systems and Signal Processing February 17, 2015

12 Introduction

13 AE testing has become a recognized nondestructive test (NDT) method,
14 commonly used to detect and locate defects in mechanically loaded structures
15 and components. AE can provide comprehensive information on the origina-
16 tion of a discontinuity (flaw) in a stressed component and also pertaining to
17 the development of this flaw as the component is subjected to continuous or
18 repetitive load [1]. Moreover, the method has been developed and applied in
19 numerous structural components, such as steam pipes and pressure vessels,
20 and in the research areas of rocks, composite materials and metals [2].

21 Acoustic emissions (AE) are stress waves produced by the sudden internal
22 stress redistribution of the materials caused by the changes within the struc-
23 ture [3]. For polymer-composite materials, these changes are mainly due to
24 crack initiation and growth, crack opening and closure, fiber breakage and
25 fiber-matrix debonding. The use of AE for structural health monitoring has
26 been investigated several decades ago with the objective to predict material
27 failure [4, 5, 6].

28 With a huge noisy amount of data originating from fatigue loading tests,
29 a major challenge in the use of AE technique is to associate each signal to a
30 specific AE source related to noise or to a damage mechanism. This analysis
31 is a non-trivial task for two main reasons. First, AE signals are complex so
32 that it has to be characterized by multiple relevant features. Second, there is
33 generally no *a priori* knowledge of the acoustic signatures of damage events
34 which are generally scattered due to the high variability of the properties of
35 composite materials [7].

36 In the literature, dealing with the challenge of massive data due to high

sensitivity of AE sensors and to long-term fatigue loading experiments, several processing approaches have been proposed [8, 9, 10]. In [8], it is considered that only signals with amplitude higher than 70 dB or recorded above 80% of peak load contain information related to damage mechanisms. In [9], “friction emission” tests in which the maximum cyclic load was decreased to a level that was insufficient to generate crack growth were performed to understand the AE signal characteristics arising from hydraulics, machine start/stop and slippage. All of the AE events at this lower peak load were therefore assumed to be due to friction emission. Emission having the characteristics of friction emission was then filtered. A more complex denoising process developed by [10] that combines Principal Component Analysis (PCA) and K-means and several validation techniques was presented to be able to classify more than 60% of the detected signals as noise during long time corrosion monitoring of a pre-damaged post tensioned concrete beam.

High dimensional feature space reduction is a remaining challenge to statistical processing and classification of AE data. In the literature, many approaches for AE data processing [1, 11] rely on the Principal Component Analysis (PCA). The PCA takes a set of features calculated from AE signals, such time-frequency features, and generates a set of artificial variables made of a linear combination of the input features depicting the largest variance. Other approaches [12, 13, 14] rely on a specific subset of features such as energy, rise time, duration, amplitude [12] or have reduced the dimension of the feature space by using complete link hierarchical clustering in order to merge the correlated features into groups [13]. Those apply a greedy approach that generates all possible feature combinations and then selects the

one which optimizes a given criterion [14, 15]. The goal of the criterion is
 generally to evaluate the quality of the partition provided by the cluster-
 ing. It can be noticed that the PCA and the K-means clustering method
 are theoretically related to each other as shown in [16]. An alternative ap-
 proach to Euclidean distance-based clustering methods was proposed [17] and
 based on the Gustafson-Kessel algorithm (GK) [18]. It makes use of a modi-
 fied Mahalanobis distance for each cluster which is iteratively adapted to fit
 ellipse-shaped clusters. The use of hyper-ellipses instead of hyper-spheres is
 more appropriate for AE clustering in presence of low density and high scat-
 tering. In the GK algorithm, the covariance between each pair of features is
 estimated so that possible redundancy or complementarity between features
 can be taken into account. The Mahalanobis distance has also been shown
 to be robust to outliers in statistical analysis [19].

The processing of large AE datasets, in particular originating from fa-
 tigue, requires to develop efficient methods in terms of memory and time
 consumption. Some approaches have been proposed which are able to work
 online (or real-time), that means that clusters parameters are updated with-
 out iterative procedure but as new data arrive. As underlined in the GK-
 based method proposed in [17] and in the Kmeans-based method developed
 in [20], external AE sources (corresponding to noise) may have an important
 influence on the clusters' updating. In this paper we propose a methodology
 to estimate efficiently the partition of AE data obtained in fatigue loading
 in presence of noise sources. The methodology also includes an automated
 sequential feature selection based on the GK algorithm and relying on quasi-
 static (QS) tests. The clusters obtained are then adapted to be applied

87 on large fatigue tests. The next section is dedicated to presentation of the
88 proposed methodology.

89 **1. Unsupervised pattern recognition**

90 The flow chart of the methodology is shown on Fig. 1.

91 [Figure 1 about here.]

92 *1.1. AE fatigue data pre-processing*

93 All acoustic emissions even originating from outside the area of interest
94 bounded by the sensors were taken into account (no spatial filtering). Thus a
95 pre-processing step of such AE data is highly important and requires adapted
96 filtering methods [21].

97 *1.1.1. Signal screening*

98 Continuous background noise due to hydraulic flows is essentially elimi-
99 nated from the AE signal by a floating signal threshold, which is adjusted
100 at a 40 dB level. This threshold makes it possible to loose signals originated
101 from friction. Optimal denoising, for instance using wavelets [22], would be
102 necessary if those signals are important for the monitoring.

103 *1.1.2. Noise model-based filtering*

104 Typical field and environmental noise such as electromagnetic interfer-
105 ence (EMI), fretting, mechanical or hydraulic vibration encountered in real
106 applications generate extraneous noise detected by the broadband and high
107 sensitive AE sensors. Assuming that this AE activity is not due to damages,

108 a noise model is built using a multivariate statistical test based on the Maha-
 109 lanobis distance as used in novelty detection [23, 24]. For that, the AE hits
 110 recorded before the loading phase are considered as representative of the AE
 111 hits corresponding to external AE sources (such as noise). The statistical
 112 mean (center of the noise model) and covariance of those samples define an
 113 ellipsoid in the feature space, and its boundary is estimated as the average
 114 of the Mahalanobis distances between each sample and the center. An AE
 115 hit recorded during loading is then considered as noise if it falls within the
 116 boundary of the ellipsoid.

117 1.2. Sequential selection algorithm of AE features

118 An automated technique is presented to detect relevant feature subsets
 119 for clustering of AE events. In contrast to feature reduction procedures (for
 120 example based on correlation dendrogram [1]) or exhaustive search of global
 121 optimal feature combinations [14], the principle of the approach is to combine
 122 gradually each feature from an available feature space with an initial feature
 123 subset [25]. The feature selection is achieved by minimizing the value of
 124 Davies and Bouldin (DB) index [26] defined by:

$$DB = \frac{1}{c} \sum_{i=1}^c \max_{i \neq j} \left\{ \frac{d_i + d_j}{D_{ij}} \right\} \quad (1)$$

125 where c is number of clusters, d_i and d_j are the average within-class distances
 126 of clusters i and j respectively, and D_{ij} denotes the distance between the
 127 two clusters i and j . This clustering validity index has been used by several
 128 authors in order to select optimal cluster number [13] or to evaluate feature
 129 subset partition [14]. The lower is its value, the better is the compactness

130 and the separability within the partition. Figure 2 shows the diagram of the
 131 proposed algorithm based on a feature filtering approach [27].

132 [Figure 2 about here.]

133 Considering an initial subset of features S (empty by default), the algorithm
 134 takes each of the available features from F to update S . This subset is then
 135 partitioned by the GK clustering algorithm. At the k^{th} iteration, a feature
 136 $f_l \in F$ is added to the current subset of features S_k , and the corresponding
 137 DB index DB_l of the partition obtained by the GK algorithm is computed.
 138 The computation of the DB index makes use of the Mahalanobis-like distance
 139 defined in the GK algorithm [18] to estimate the distance between AE hits
 140 and cluster centers and finally obtain the estimate of the average within-class
 141 distances used in Eq. 1 (d_i and d_j).

142 The subset of features S_{k+1} for the next iteration is given by $S_k \cup f_{l^*}$ with
 143 $l^* = \arg \min_l DB_l$ and the partition is then evaluated by the DB criterion.
 144 The feature that minimizes the value of DB index is selected and transferred
 145 from F to S . At each iteration, the procedure generates $|F|$ new subsets
 146 since each new subset contains the features from S plus a new one taken
 147 from the remaining features in F . The algorithm stops when no new subsets
 148 can improve the DB criterion.

149 For each iteration k , an improvement rate $IR(k)$ is calculated as follows:

$$IR(k) = \frac{DB(S_k) - DB(S_{k-1})}{DB(S_{k-1})} \quad (2)$$

150 where $DB(S_k)$ and $DB(S_{k-1})$ represent the value of the DB-index of the best
 151 feature selection for the k^{th} and $(k-1)^{th}$ iteration respectively. The sign of IR
 152 indicates if the DB criterion is improved (negative) or not (positive). For the

153 last iteration k^{last} (for which $IR(k^{\text{last}}) > 0$), if $IR(k^{\text{last}}) < \min_{k < k^{\text{last}}} |IR(k)|$
154 then the feature with the best DB-index is added to S to establish the final
155 selected feature set.

156 1.3. AE source clustering

157 Quasi-static (QS) tests are first applied to obtain a relatively low amount
158 of data compared to fatigue and by supposing that damage sources in QS
159 tests are mostly similar to fatigue. The GK algorithm is thus applied to
160 estimate the parameters of a given set of k clusters on AE originated from
161 QS tests. To cope with possibly additional AE sources that can occur during
162 fatigue [28], an additional $k + 1^{\text{th}}$ cluster is estimated based on fatigue data
163 to include all feature vectors located “far” from the previous k clusters. For
164 that, the boundary of each cluster characterized on QS tests is estimated by
165 the average of the Mahalanobis-like distance (used in GK) [24]. A feature
166 vector obtained during fatigue belongs to the $k + 1^{\text{th}}$ cluster if its distance to
167 nearest cluster is above the corresponding radius. This adaptation of clusters
168 is supposed to take into account one (or more) AE sources that is (or are)
169 not present in quasi-static tests (e.g. noise due to repeated tensile loading,
170 acoustic waves related to cumulated damage ...).

171 2. Experiments

172 Composite split disks were considered subjected to cyclic fatigue loading
173 up to failure determined when a complete break of the specimen was ob-
174 served in the hoop direction. The specimens were cyclically tested under a
175 tensile/tensile sinusoidal loading with constant amplitude and frequency of 5
176 Hz and under constant stress ratio $R = 0.1$ at room temperature. Quasi-static

177 tests were preliminarily conducted on five different specimens with a constant
 178 loading rate of 0.3 kN.s^{-1} . The static failure stress was equal to 1520 ± 165
 179 MPa. The tests were performed according to ASTM D2290 "Apparent hoop
 180 tensile strength of plastic or reinforced plastic pipe by split disk method".
 181 Rings were produced by cutting and machining filament-wound carbon fiber
 182 reinforced epoxy tubular structures intended for the manufacturing of fly-
 183 wheel rotors with a $(90^\circ)_6$ lay-up configuration. The transient elastic waves
 184 were recorded during test at the material surface using a multi-channels data
 185 acquisition system from EPA (Euro Physical Acoustics) corporation (MIS-
 186 TRAS Group). The system is made up of miniature piezoelectric sensors
 187 (micro-80) with a range of resonance of 250 - 325 kHz, preamplifiers with a
 188 gain of 40dB and a 20 - 1000 kHz filter, a PCI card with a sampling rate
 189 of 1MHz and the AEWIn software. Two AE sensors were coupled on the
 190 specimen faces using silicon grease. The experimental set-up is shown in
 191 Fig. 3.

192 [Figure 3 about here.]

193 The calibration of the system was performed after installation of the trans-
 194 ducers on the specimen and before each test using a pencil lead break pro-
 195 cedure. A part of the ambient noise was filtered using a threshold of 40dB.
 196 The acquisition parameters: PDT (Peak Definition Time) = $60 \mu\text{sec}$; HDT
 197 (Hit Definition Time) = $120 \mu\text{sec}$ and HLT (Hit Lock Time) = $300 \mu\text{sec}$ were
 198 optimized for this specific experimental configuration to extract transient sig-
 199 nals. The optimization of these time-driven parameters was performed using
 200 the standard pencil-lead breakage proposed by Hsu and Nielsen [29]. Many
 201 features such as absolute energy, counts, hits, amplitude, duration, frequency

centroid were calculated from recorded waves.

3. Results and discussion

According to different percentages of the ultimate tensile stress determined in the tensile test (1520 MPa), the S-N curve was obtained as illustrated in Figure 4.

[Figure 4 about here.]

Nine samples were used to generate the S-N curve. This was a good compromise between the six specimens recommended by ASTM D-3479 for preliminary and exploratory test campaign and the twelve specimens required for research and development on testing of components and structures. The results presented in this work are part of a wider study including the generation of S-N curves of different types of composites (with different carbon fibers) and with different lay-up configurations. The main goal is to select a composite of choice for the application concerned, namely rotors of flywheels.

Four datasets were considered denoted as A1 (quasi-static test) and A2, A3 and A4 (fatigue tests for 90%, 80% and 70% of the ultimate tensile strength respectively). A brief description of the obtained datasets is summarised in Table 1.

[Table 1 about here.]

3.1. Noise reduction

According to the scenario of the quasi-static test A1 (Fig. 5(a)), around the time-instant t_1 , the actuator was pressurized and the stress was applied

224 only at t2. The noise modeling phase (Section 1.1.2) has been made from AE
225 data recorded before t1, i.e. while the specimen was let in its environment
226 without any mechanical loading. Noise during loading is then filtered by this
227 model.

228 [Figure 5 about here.]

229 Figure 5(b) and 5(c) represent dataset A1 (made of 52,832 AE hits) in the
230 duration-amplitude space segmented into three populations: noise before and
231 during loading in Figure 5(b), and denoised data after application of noise
232 model in Figure 5(c). The two first populations (noise) possess the same
233 characteristics, the same location and the same scattering. This observation
234 is justified by the graphic of AE cumulated energy in Fig. 6(a). Indeed, the
235 level of AE cumulated energy of noise before and during loading is negligi-
236 ble and the total energy is conserved within denoised data while the latter
237 occupies only 12% of the whole dataset in terms of quantity (Fig. 6(b)).

238 [Figure 6 about here.]

239 The application of the noise model to fatigue dataset A3 made of
240 1,682,434 AE hits led to a similar separation between noise and denoised
241 data (Fig. 7(a)). In spite of 93% of AE hits recorded associated to “noise”
242 (Fig. 7(c)), this highest population represents negligible AE cumulated en-
243 ergy level in comparison with that of denoised data (Fig. 7(b)).

244 [Figure 7 about here.]

245 3.2. Feature selection

246 Many energy-based approaches of damage characterization or identifica-
247 tion have been studied since AE energy provides a good correlation with
248 damage mechanisms. Thus, in this work, absolute energy (Fig. 8) is used
249 to initialize the subset of relevant features. As the number of clusters is un-
250 known, 3 cases were addressed to check the stability of the selection algorithm
251 by considering 4, 5 and 6 clusters.

252 The selection algorithm was applied on the quasi-static dataset A1 with 4
253 clusters. At the first iteration, given the absolute energy feature, the optimal
254 DB index is given by the combination with the amplitude feature (Fig. 8(a)).
255 At the second iteration, the best score was obtained by the combination with
256 the MARSE energy (Fig. 8(b)). No more improvement of the DB index is
257 made at the next iteration, so the algorithm is stopped by selecting the subset
258 made of absolute energy, amplitude and MARSE energy. The same selection
259 result was obtained with 5 and 6 clusters. In what follows, 4 clusters are
260 used as initial number of AE sources.

261 [Figure 8 about here.]

262 3.3. AE source classification

263 3.3.1. Sequence of AE hits in the quasi-static case

264 The denoised and selected feature subset obtained previously is now used
265 to identify the clusters in quasi-static dataset A1 using the GK clustering
266 algorithm. Four well-separated clusters with different sizes and shapes have
267 been obtained in the duration-amplitude space (Fig. 9(a)). After projection
268 onto the amplitude dimension, four distinct distributions can be obtained,

among which three are located above 75 dB. These distributions have been often used to identify AE sources [30, 31].

Ono and Gallego [2] recently underlined a misconception that fiber fracture always produces high-energy event, and that still persists to this day. For the considered material, the damage process involves fiber *tow* breakage. If the breakage of an elementary fiber ($7\mu m$ diameter) can cause the release of low energy transient, the breakage of fiber tows including hundreds or thousands of elementary fibers (up to 12,000 in the considered material) are likely to induce highly energetic signals.

As a complementarity view, the temporal evolution of the logarithm of the Cumulated Sum of Cluster Appearance (logCSCA) [17] has been depicted in Fig. 9(c) (for each cluster) together with the cumulated energy and the load. When an AE hit (emitted after the activation of an AE source) is associated to a given cluster at a given time, the corresponding logCSCA curve depicts a *step*. When several consecutive steps appear in a short time period, this visualisation allows to point out that the activity of the corresponding AE source is particularly sustained which may be related to propagations of cracks [32].

In the sequence shown in Fig. 9(c), the first cluster is activated at the very beginning before applying the load. Despite the number of AE hits in this cluster is important (73% at the end), the cumulated energy of AE hits in this cluster is the lowest one among all clusters (Fig. 9(e)). These observations are coherent with the activation of an AE source related to mechanical and hydraulic emission such as vibration and friction between the specimen and the half-cylinders.

Both cluster 2 and 3 start early when the actuator has been pressurized. The main activity of cluster 2 occurs after a certain level of load (Fig. 9(c)) and the cumulated energy of AE hits in this cluster (Fig. 9(e)) as well as the amplitudes (> 95 dB, Fig. 9(a)) are the highest ones compared to all other clusters. The number of AE hits in this cluster is particularly important at the end of the test, as expected with the ruine of the specimen induced by a cascade of fiber **tow** breakage. This cluster is thus related to the activity of highly energetic sources, in particular carbon fiber **tow** breakage.

The low cumulated energy in cluster 3 as well as the amplitudes around 75 and 90 dB make this cluster related to minor damage (probably matrix micro-cracks).

The partition also emphasizes an important cascade of AE hits at $t \approx 352$ s during which the activity of cluster 3 increases importantly and this increase is synchronised with both the appearance of cluster 4 and a high activity of cluster 3. The load level at this time, the mean value of amplitudes in cluster 4 (around 95 dB, Fig. 9(a)) and the level of the cumulated energy in this cluster (around 13% of the total cumulated energy, Fig. 9(e)) make this cluster related to macro-cracking and interface failures starting around the specimen's notches and propagating gradually in the hoop direction.

3.3.2. Sequence of AE hits in a fatigue test

Afterwards, the model estimated on A1's AE hits is used to infer the partition on the fatigue dataset A3. Direct application of the model generates overlapping zones between clusters in the duration-amplitude space of A3 (Fig. 9(b)). **We can observe a similar distribution of clusters in this feature space compared to A1 (Fig. 9(a)). However, we can also observe clusters**

319 overlap, particularly important between clusters 2 and 3. As a consequence,
 320 the projection onto the amplitude axis would not give distinct distributions
 321 as for the quasi-static test. This phenomenon finds its origins in the fact
 322 that, compared to quasi-static tests, additional mechanisms can play a role
 323 during fatigue such as the temperature [33] or the cycling which implies
 324 crack opening/closing initially not observed during QS tests [28]. Therefore,
 325 it was expected to find out that a pattern recognition model learned from a
 326 quasi-static test and simply applied on a fatigue test may present a limited
 327 generalization capability. Based on the assumption that a new AE source
 328 is activated during fatigue and which has not been observed in quasi-static
 329 tests, the proposed methodology (Section 1.3) includes the creation of new
 330 cluster to cope with this problem. The result is a new segmentation with less
 331 overlapping between clusters as shown in Fig. 9(d).

332 [Figure 9 about here.]

333 The comparison of partitions with the previous quasi-static test yield
 334 similar conclusions concerning the possible damage scenario. The main dif-
 335 ference holds in the position of the new cluster, which has been automatically
 336 found from AE hits. Indeed, the cluster 3 identified as the friction and pos-
 337 sibly micro-cracking in the quasi-static test (Fig. 9(a)) was split (AE sources
 338 3a and 3b, Fig. 9(b)). The signatures of AE hits in both clusters in terms of
 339 amplitudes, durations and energies (Fig. 9(f)) are quite different despite the
 340 fact that the clusters are pretty close in the duration-amplitude space. The
 341 evolution of AE cumulated energy of each source (Fig. 9(f)) brings useful in-
 342 terpretations about the damaging process during fatigue. Despite its smallest

343 population, AE source 2 is dominant in term of energy at the end of test as
344 for the quasi-static test and is associated to severe damage mechanisms re-
345 lated to carbon fibers. AE source 1 is the most scattered and populated but
346 represents negligible contribution compared to the total energy. As for the
347 quasi-static case, this cluster may represent the activation of an AE source
348 related to mechanical and hydraulic systems [34]. AE source 4 generates AE
349 hits with the longest duration and the highest energy that may be related to
350 macro-cracking and interface failures.

351 [Figure 10 about here.]

352 [Figure 11 about here.]

353 Figure 11 represents the positioning of clusters onto the load level for
354 the fatigue dataset A3. This figure enables one to visualise the load level
355 when the AE sources are activated. On dataset A3, it can be observed in
356 Fig. 10(a) to 10(f) that during 20%-fatigue-life of the specimen, many AE
357 hits appear, related to all AE sources. This phenomenon is well known as
358 the accommodation phase [35] which generally appears at the first stage
359 of materials undergoing fatigue testing and may lead to partial fractures.
360 Indeed, AE hits with high energy (from AE sources 2 and 4) are activated
361 during this phase (and during failure). After this stage, the clusters' activities
362 globally slow down for a while (stabilization phase). Beyond 65% of the
363 fatigue life, an important number of highly energetic AE hits occur up to the
364 ruine of the specimen (from AE sources 2 and 4). It is interesting to notice
365 a repetitive phenomena that takes place all along the test and represented
366 by the activation of AE sources 3a and 3b: The latter is mainly activated in

loading phases while the former occurs in unloading phase (Fig. 11(b)). As for the quasi-static test, the latter may correspond to internal frictions and interfaces fretting as observed in previous papers [36]. It can also be observed that the AE hits originated from these clusters occur between 5-7 kN at the beginning of the tests and between 3-4 kN at the end. AE source 3a is much more activated than AE source 3b between 20% and 50%, then the activity of 3b substantially increases until the ruine. This increasing is followed by the activation of AE source 5 that is particularly active between 70% and 90%, just before the ruine. Therefore, as expected, the fatigue plays a role on the loading level required to activate some sources and the chronology of activation may give insights to the understanding of damage mechanisms.

3.3.3. Sequence of AE hits in two other fatigue tests

The complexity of damage mechanisms involved during fatigue is illustrated in this section. For that, two other specimens denoted as A2 and A4, corresponding to 90% and 70% of the tensile strength respectively, are considered. The behavior of A2 is similar to the previous specimen A3 as depicted in Fig. 12(a). The activity of the AE hits including high energy and high duration signals is rather high (relatively to the remaining AE hits) at the very beginning of loading and increases again at about 60% of the specimen life, as for A3. Although AE hits generated by AE source 3a are more scattered than in the previous test, overlaps between clusters related to this source and to AE source 3b have also been detected by the proposed algorithm. Table 2 summarises the clusters assigned to each AE source according to the previous observations.

[Table 2 about here.]

Rather different than the previous tests, the partition obtained on dataset A4 at 70% of the ultimate static strength is depicted in Fig. 12(b). The initial (accommodation) phase occurs within the first cycles as for the previous loading levels, but it is then followed by a silence of most of AE sources. Only AE source 1 is activated (representing possible external sources which has been filtered out) and a few highly energetic AE hits occur (such as fiber tow breakage). Then, at 20% of the fatigue life, a progressive activation of all AE sources can be observed. In the load band 2 – 10 kN, only cluster 1 is activated but this band is gradually reduced with respect to the number of cycles to reach 4 – 7 kN when approaching the end-of-life. The progressive and continuous reduction of the band beyond which clusters are activated can be of interest for predicting the remaining lifetime of the composite if confirmed on other specimens and lower loading levels. It can also be noticed that more AE hits related to AE sources 2 and 4 (i.e. with the highest energy) can be found compared to the two previous specimens. Therefore, the failure process of specimen A4 tested at 70% of the ultimate tensile strength is more gradual and more related to the progressive weakness of the material during the repeated stress until the ruine.

[Figure 12 about here.]

Conclusion

An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer-composite materials has been presented to

414 tackle different existing challenges of AE analysis and damage detection: 1)
 415 data pre-processing, especially noise reduction; 2) automatic and fast fea-
 416 ture selection; 3) clustering of massive data from fatigue tests with cluster
 417 adaptation. The methodology relies on the estimation of clusters during
 418 static tests. Its application to big fatigue data based on the adaptation phase
 419 allows to add a new cluster to cope with new AE sources. The assignment
 420 of a cluster to a AE hit is not iterative and only requires to find the closest
 421 cluster by using a Mahalanobis-like distance that allows to cope with data
 422 scattering. The processing of a fatigue dataset is made faster than itera-
 423 tive procedures which requires to load a dataset and to perform iterative
 424 optimization on large matrices.

425 The first results on three real fatigue tests of thermoset ring-shaped CFRP
 426 involving until 10 millions AE hits demonstrate that the proposed method-
 427 ology allows to identify some relevant clusters. Of particular interest:

- 428 • Four main phases have been identified: Accommodation with many AE
 429 hits with the highest energy and amplitude (0-20% of the lifetime), a
 430 slowdown of AE activity (20-50%), a resumption of the AE activity
 431 (50-85%) and a failure progress up to the final failure (85-100%). The
 432 fatigue at 70% of the ultimate strength depicts a particular pattern
 433 during the degradation involving an envelop which gradually reduces
 434 until the ruine.
- 435 • Two clusters detected by the adaptation phase occur at similar loading
 436 levels. A modification of their kinetics with report to the cumulated
 437 loading lets suppose that those two clusters can be due to damage. It
 438 is also interesting to emphasize that the level required to activate the

439 AE sources related to those two clusters depicts a slight and progressive
440 decreasing together with the degradation of the material until the ruine.

441 The visualization of clusters in the amplitude-duration feature, the loga-
442 rithm of the cumulated AE hits and energy in each cluster as well as the
443 the positioning of clusters onto the loading level have allowed to connect
444 some clusters to possible AE sources. In order to validate the identification
445 of AE sources observed, complementary non-destructive techniques and in-
446 situ measurements is under study on more specimens. The application of
447 the proposed methodology is currently investigated on thermoplastic CFRP
448 composites and compared to finite element models [37]. Finally, the pro-
449 posed methodology is under improvement for robust AE-based prognostics
450 of composite structures.

451 **Acknowledgement**

452 This work has been carried out in the framework of the Laboratory of
453 Excellence ACTION through the program "Investments for the future" man-
454 aged by the National Agency for Research (references ANR-11-LABX-01-01).
455 The authors thank the reviewers for the relevant and fruitful comments.

456 [1] A. A. Anastassopoulos, T. P. Philippidis, Clustering methodology for
457 the evaluation of ae from composites, J. of Acoust. Emission 13 (1-2)
458 (1995) 11-22.

459 [2] K. Ono, A. Gallego, Research and applications of ae on advanced com-
460 posites, Journal of Acoustic Emission.

- 461 [3] M. Huang, L. Jiang, P. Liaw, C. Brooks, R. Seeley, D. Klarstrom, Using
462 acoustic emission in fatigue and fracture materials research, *The Journal*
463 *of The Minerals, Metals and Materials Society* 50 (1998) 1–12.
- 464 [4] L. J. Graham, Characterization of acoustic emission signals and ap-
465 plication to composite structures monitoring, in: *Proceedings of the*
466 *ARPA/AFML Review of Progress in Quantitative NDE*, 1975.
- 467 [5] C. Murthy, B. Dattaguru, A. Rao, Application of pattern recognition
468 concepts to acoustic emission signals analysis, *Journal of Acoustic Emis-*
469 *sion* 6 (1987) 19–28.
- 470 [6] M. Otsu, K. Ono, Pattern recognition analysis of acoustic emission from
471 unidirectional carbon fiber-epoxy composites by using autoregressive
472 modeling, *Journal of Acoustic Emission* 6 (1) (1987) 61–70.
- 473 [7] P. Pineau, F. Dau, Subsampling and homogenization to investigate vari-
474 ability of composite material mechanical properties, *Computer Methods*
475 *in Applied Mechanics and Engineering* 241244 (2012) 238 – 245.
- 476 [8] J. Henry, Z. Aboura, K. Khellil, S. Otin, Suivi de l’endommagement en
477 fatigue d’un composite a renfort interlock carbone/epoxy par emission
478 acoustique, *Mater. & Tech.* 100 (6-7) (2012) 643–652.
- 479 [9] J. Yu, P. Ziehl, B. Zarate, J. Caicedo, L. Yu, V. Giurgiutiu, B. Metro-
480 vich, F. Matta, Quantification of fatigue cracking in CT specimens with
481 passive and active piezoelectric sensing, in: *Proc. SPIE*, Vol. 7649, 2010,
482 pp. 76490R–76490R–12.

- 483 [10] L. Calabrese, G. Campanella, E. Proverbio, Noise removal by cluster
484 analysis after long time AE corrosion monitoring of steel reinforcement
485 in concrete, *Constr. Build Mater.* 34 (2012) 362–371.
- 486 [11] N. Godin, S. Huguet, R. Gaertner, L. Salmon, Clustering of ae signals
487 collected during tensile tests on unidirectional glass/polyester composite
488 using supervised and unsupervised classifiers, *Ndt&e Int* 37 (4) (2004)
489 253–264.
- 490 [12] R. Gutkin, C. Green, S. Vangrattanachai, S. Pinho, P. Robinson, P. Cur-
491 tis, On ae for failure investigation in CFRP: pattern recognition and
492 peak frequency analyses, *Mech. Syst. Signal Pr.* 25 (4) (2011) 1393–
493 1407.
- 494 [13] M. Moevus, N. Godin, M. RMili, D. Rouby, P. Reynaud, G. Fantozzi,
495 G. Farizy, Analysis of damage mechanisms and associated ae in two
496 SiCf/[SiBC] composites exhibiting different tensile behaviours. Part II:
497 unsupervised ae data clustering, *Compos. Sci. Tech.* 68 (6) (2008) 1258–
498 1265.
- 499 [14] M. Sause, A. Gribov, A. Unwin, S. Horn, Pattern recognition approach
500 to identify natural clusters of ae signals, *Pattern Recogn. Lett.* 33 (1)
501 (2012) 17–23.
- 502 [15] M. Halkidi, Y. Batistakis, M. Vazirgiannis, On clustering validation
503 techniques, *J. Intell. Inf. Syst.* 17 (2-3) (2001) 107–145.
- 504 [16] C. Ding, X. He, K-means clustering via principal component analysis,
505 in: *Proc. ICML, Banff, Canada*, 2004.

- 506 [17] V. Placet, E. Ramasso, L. Boubakar, N. Zerhouni, Online segmentation
507 of ae data streams for detection of damages in composite structures in
508 unconstrained environments, in: Int. Conf. Struct. Saf. Reliab., New
509 York, 2013.
- 510 [18] D. Gustafson, W. Kessel, Fuzzy clustering with a fuzzy covariance ma-
511 trix, in: IEEE Int. Conf. Decis. Control, Vol. 17, 1978, pp. 761–766.
- 512 [19] D. Peel, G. J. McLachlan, Robust mixture modelling using the t distri-
513 bution, *Statistics and Computing* 10 (2000) 339–348.
- 514 [20] E. Pomponi, A. Vinogradov, A real-time approach to acoustic emission
515 clustering, *Mechanical Systems and Signal Processing* 40 (2) (2013) 791
516 – 804.
- 517 [21] V. Barat, Y. Borodin, A. Kuzmin, Detection of AE signals against back-
518 ground friction, *J. of Acoust. Emission* 29.
- 519 [22] D. Donoho, De-noising by soft-thresholding, *IEEE Transactions on In-*
520 *formation Theory* 41 (3) (1995) 613–627.
- 521 [23] C. Farrar, K. Worden, Unsupervised learning novelty detection, in:
522 *Struct. Health Monit.*, 2012, pp. 321–360.
- 523 [24] L. Serir, E. Ramasso, P. Nectoux, N. Zerhouni, E2GKpro: An evidential
524 evolving multi-modeling approach for system behavior prediction with
525 applications, *Mechanical Systems and Signal Processing* 37 (1-2) (2013)
526 213–228.

- 527 [25] S. Guerif, Unsupervised variable selection: when random rankings sound
528 as irrelevancy, *JMLR Proc.* 4 (2008) 163–177.
- 529 [26] D. Davies, D. Bouldin, A cluster separation measure, *IEEE T. Pattern*
530 *Anal. PAMI-1* (2) (1979) 224 –227.
- 531 [27] G. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selec-
532 tion problem, *Mach. Learn.: Proc. Int. Conf.* (1994) 121–129.
- 533 [28] I. D. Baere, W. V. Paepegem, J. Degrieck, Assessment of the inter-
534 laminar behaviour of a carbon fabric reinforced thermoplastic lap shear
535 specimen under quasi-static and tension-tension fatigue loading, *Poly-*
536 *mer Testing* 32 (7) (2013) 1273 – 1282.
- 537 [29] N. Hsu, F. Breckenridge, Characterization and calibration of acoustic
538 emission sensors, *Materials Evaluation* 39 (1) (1981) 60–68.
- 539 [30] L. J. Graham, R. K. Elsley, Characteristics of ae signals from composites,
540 in: *ARPA/AFML Rev. Prog. Quant. NDE*, 1977, pp. 219–225.
- 541 [31] D. Short, J. Summerscales, Amplitude distribution ae signatures of uni-
542 directional fibre composite hybrid materials, *Composites* 15 (3) (1984)
543 200–206.
- 544 [32] D. Sornette, Statistical physics of rupture in heterogeneous media, in:
545 *Handb. Mater. Model.*, 2005, pp. 1313–1331.
- 546 [33] J. R. Gregory, S. M. Spearing, Constituent and composite quasi-static
547 and fatigue fracture experiments, *Composites Part A: Applied Science*
548 *and Manufacturing* 36 (5) (2005) 665 – 674.

- 549 [34] M. Wevers, G. V. dijk, W. Desadeleer, M. Windelmans, K. V. D.
550 Abeele, Acoustic emission for on-line monitoring of damage in various
551 application fields, *Journal of Acoustic emission* 22.
- 552 [35] G. Degallaix, A. Seddouki, S. Degallaix, Low cycle fatigue of a duplex
553 stainless steel alloyed with nitrogen, in: *Proc. 3rd Int. Conf. Low Cycle*
554 *Fatigue Elasto-Plast. Behav. Mater.*, Vol. 3, Berlin, 1992.
- 555 [36] J. Cheng, H.-J. Li, S.-Y. Zhang, L.-Z. Xue, W.-F. Luo, W. Li, Inter-
556 nal friction behavior of unidirectional carbon/carbon composites after
557 different fatigue cycles, *Mater. Sci. Eng.: A* 600 (2014) 129–134.
- 558 [37] M. G. Sause, S. Richler, Finite element modelling of cracks as acoustic
559 emission sources, *Journal of Nondestructive Evaluation* 34 (1).

560 List of Figures

561	1	Unsupervised damage detection methodology	27
562	2	Sequential feature selection diagram	28
563	3	Experimental set-up for tensile test on split-disk specimen. (1)	
564		fixture, (2) notched ring specimen, (3) half-cylinder, (4) AE	
565		sensor, (5) notched region, (6) load.	29
566	4	S-N curve of all tested specimens with ●: 90% of the ultimate	
567		tensile strength, ▼: 80%, ◆: 70% and ■: 60%.	30
568	5	Quasi-static dataset A1: (a) Loading profile; (b) and (c) Du-	
569		ration vs. Amplitude for AE hits detected as noise (remaining	
570		data surimposed in light gray) and for for denoised data re-	
571		spectively	31
572	6	Quasi-static dataset A1: (a) AE cumulated energy; (b) Per-	
573		centage in terms of population	32
574	7	Fatigue dataset A3: (a) Duration vs. Amplitude; (b) AE cu-	
575		mulated energy; (c) Percentage in terms of population	33
576	8	Case of 4 clusters: (a) first selection giving amplitude feature	
577		as the best; (b) second selection giving feature MARSE energy	
578		as the best.	34
579	9	Left – Clustering result on dataset A1 (quasi-static): (a) Par-	
580		tition in the Duration vs. Amplitude space; (c) Evolution of	
581		the cumulated number of hits in each cluster (log CSCA); (e)	
582		Cumulated energy of each source. Right – Testing phase on	
583		dataset A3 (fatigue): (b) Direct classification without adapta-	
584		tion; (d) Adaptive classification; (f) Cumulated energy of each	
585		AE source.	35
586	10	Classified AE events during cyclic loading of specimen A3	
587		(80%): (a) All sources during the whole test; (b)-(f) Individual	
588		AE source	36
589	11	AE events during cyclic loading of specimen A3 (80%): Close-	
590		up view (a) at the beginning and (b) at the end of the test. . .	37
591	12	Visualization of classified AE events during cyclic loading A2	
592		(90% of the ultimate strength) and A4 (70%)	38

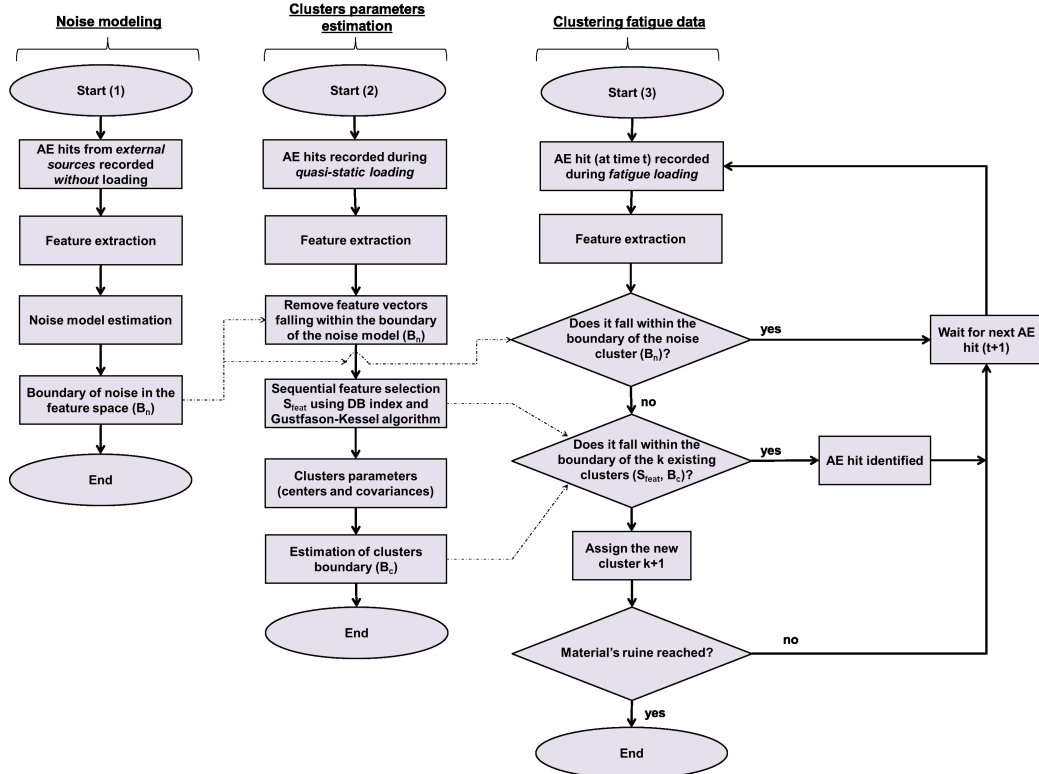


Figure 1: Unsupervised damage detection methodology

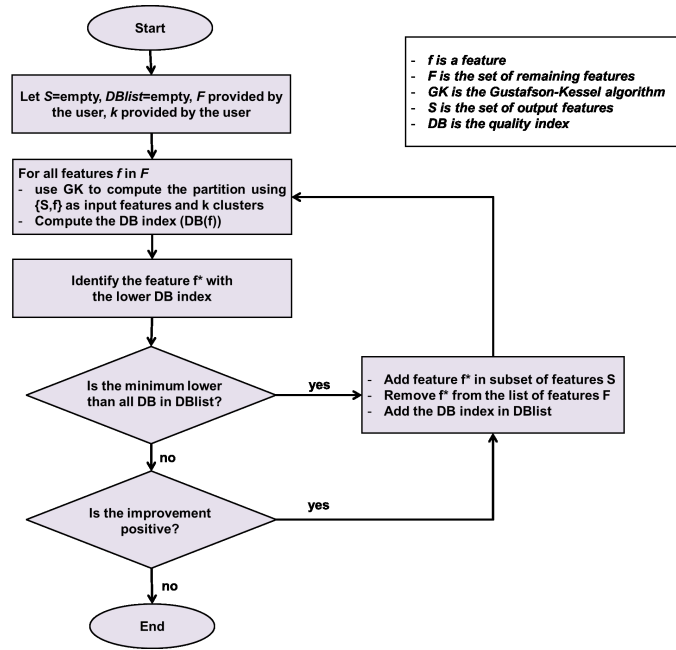


Figure 2: Sequential feature selection diagram

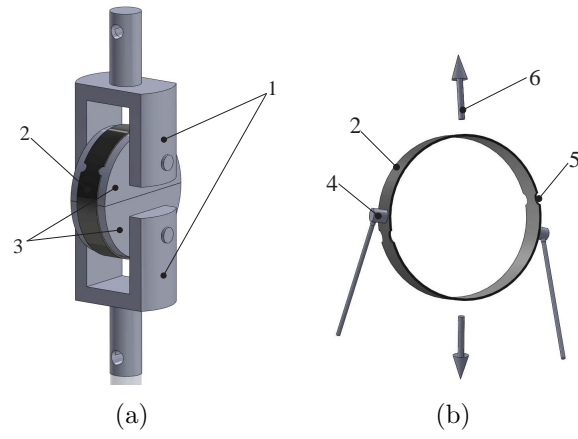


Figure 3: Experimental set-up for tensile test on split-disk specimen. (1) fixture, (2) notched ring specimen, (3) half-cylinder, (4) AE sensor, (5) notched region, (6) load.

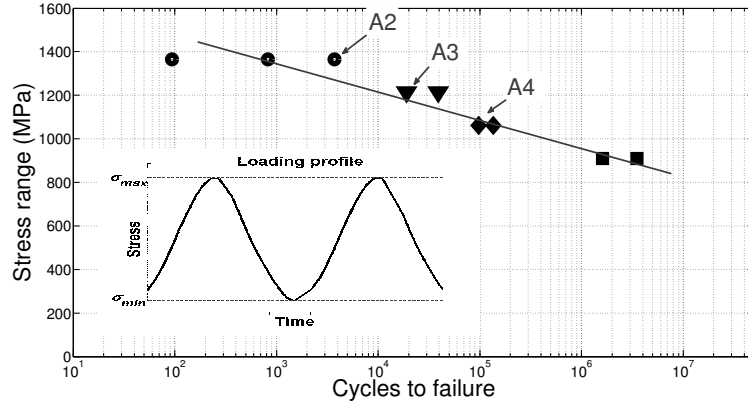
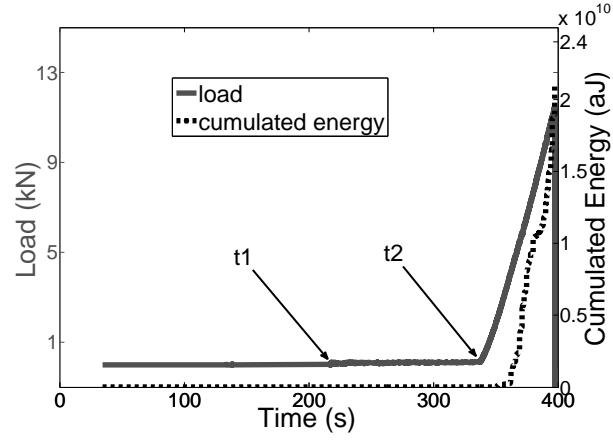
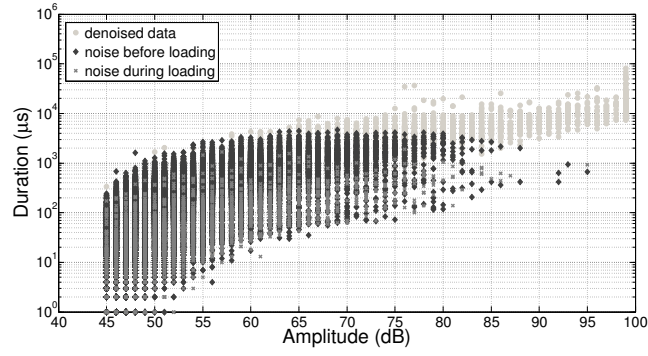


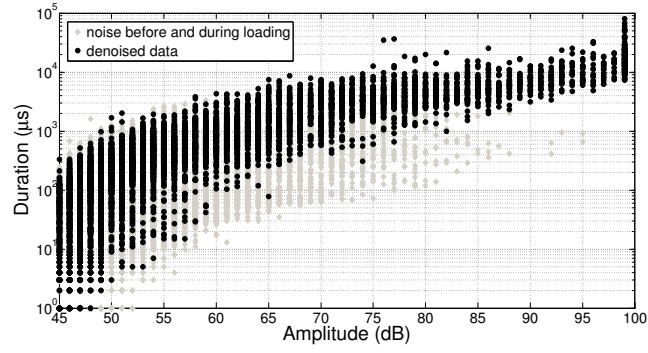
Figure 4: S-N curve of all tested specimens with ●: 90% of the ultimate tensile strength, ▼: 80%, ◆: 70% and ■: 60%.



(a)



(b)



(c)

Figure 5: Quasi-static dataset A1: (a) Loading profile; (b) and (c) Duration vs. Amplitude for AE hits detected as noise (remaining data surimposed in light gray) and for for denoised data respectively

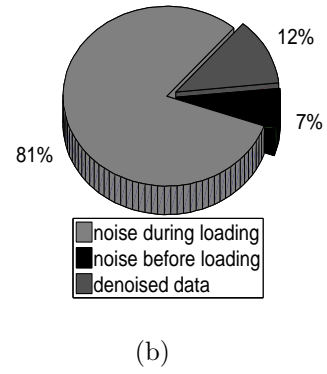
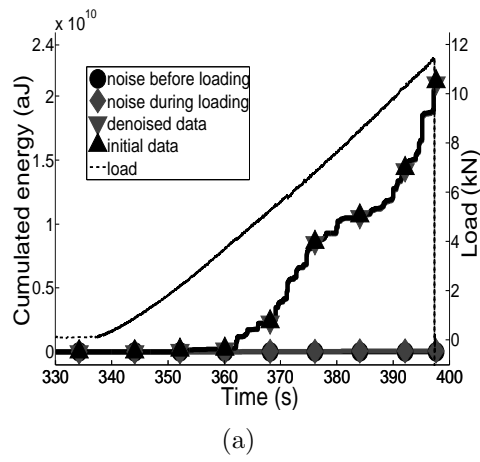


Figure 6: Quasi-static dataset A1: (a) AE cumulated energy; (b) Percentage in terms of population

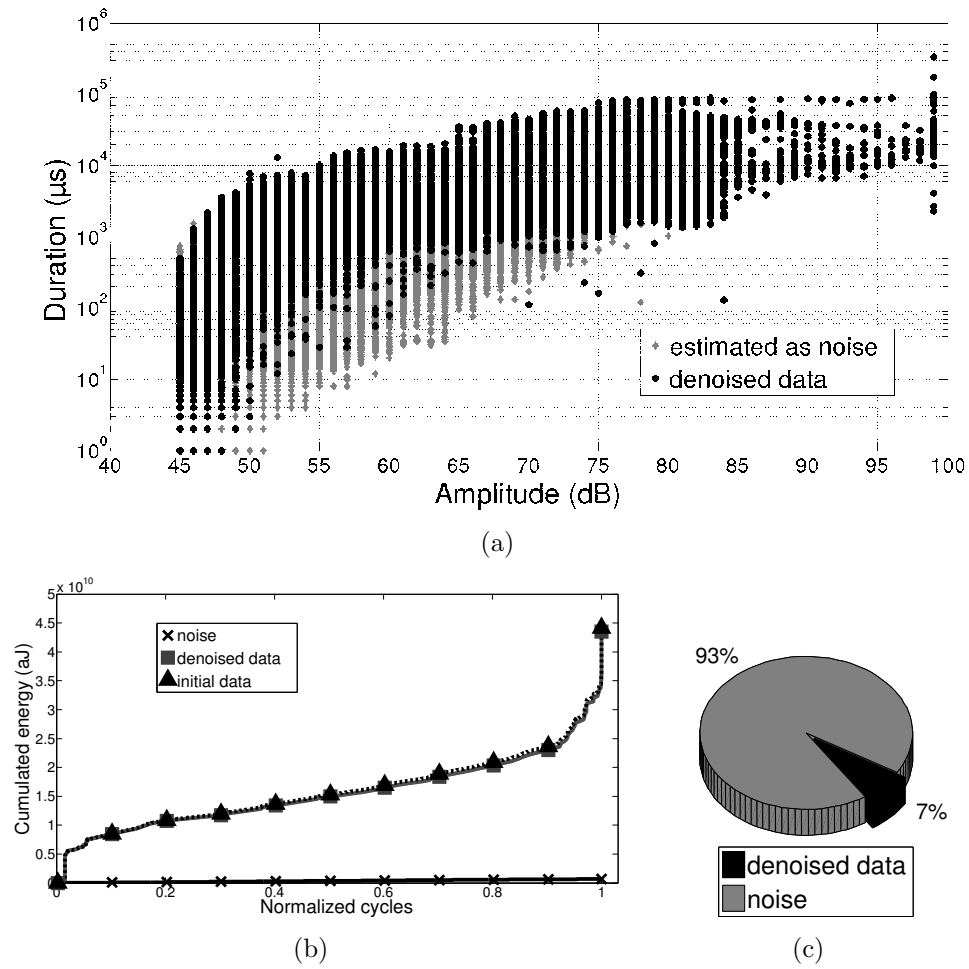
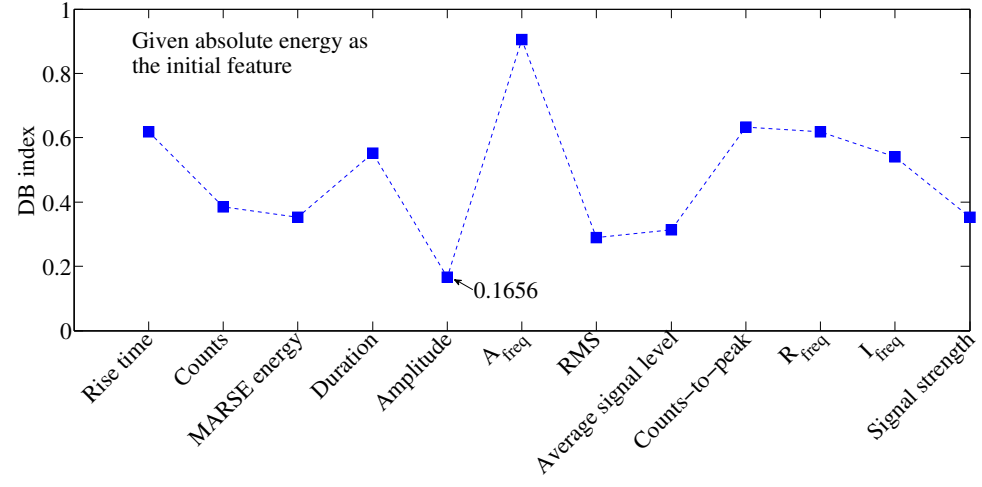
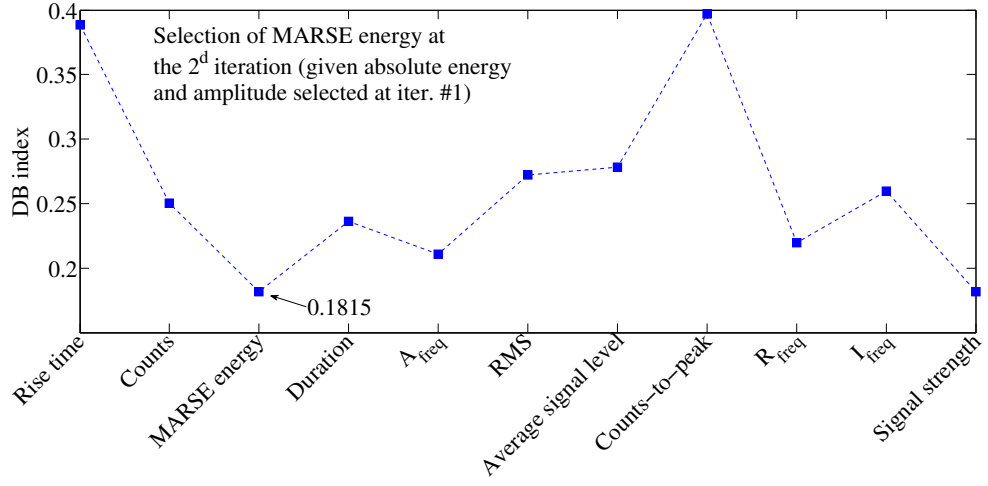


Figure 7: Fatigue dataset A3: (a) Duration vs. Amplitude; (b) AE cumulated energy; (c) Percentage in terms of population



(a)



(b)

Figure 8: Case of 4 clusters: (a) first selection giving amplitude feature as the best; (b) second selection giving feature MARSE energy as the best.

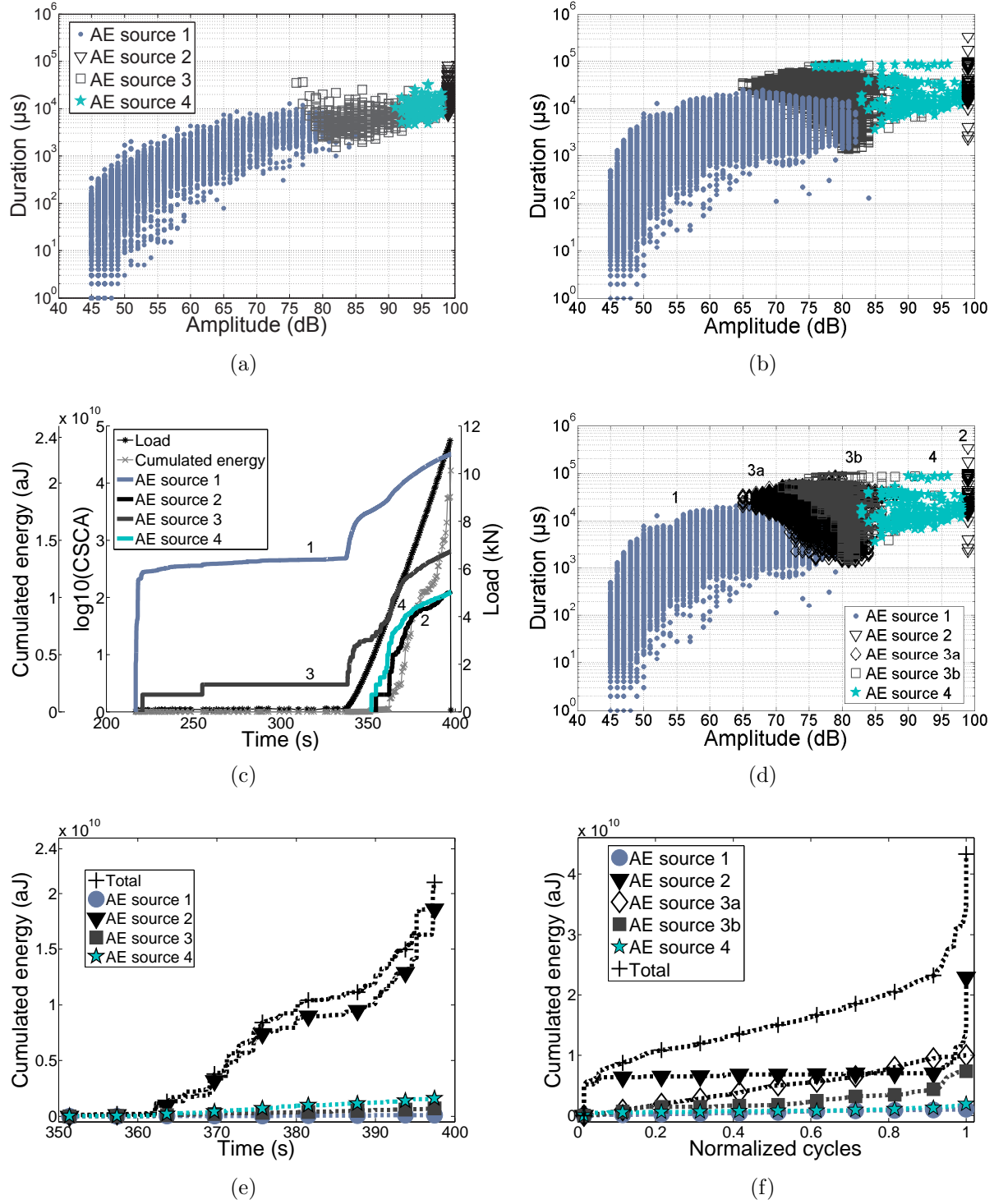


Figure 9: **Left** – Clustering result on dataset A1 (quasi-static): (a) Partition in the Duration vs. Amplitude space; (c) Evolution of the cumulated number of hits in each cluster (log CSCA); (e) Cumulated energy of each source. **Right** – Testing phase on dataset A3 (fatigue): (b) Direct classification without adaptation; (d) Adaptive classification; (f) Cumulated energy of each AE source.

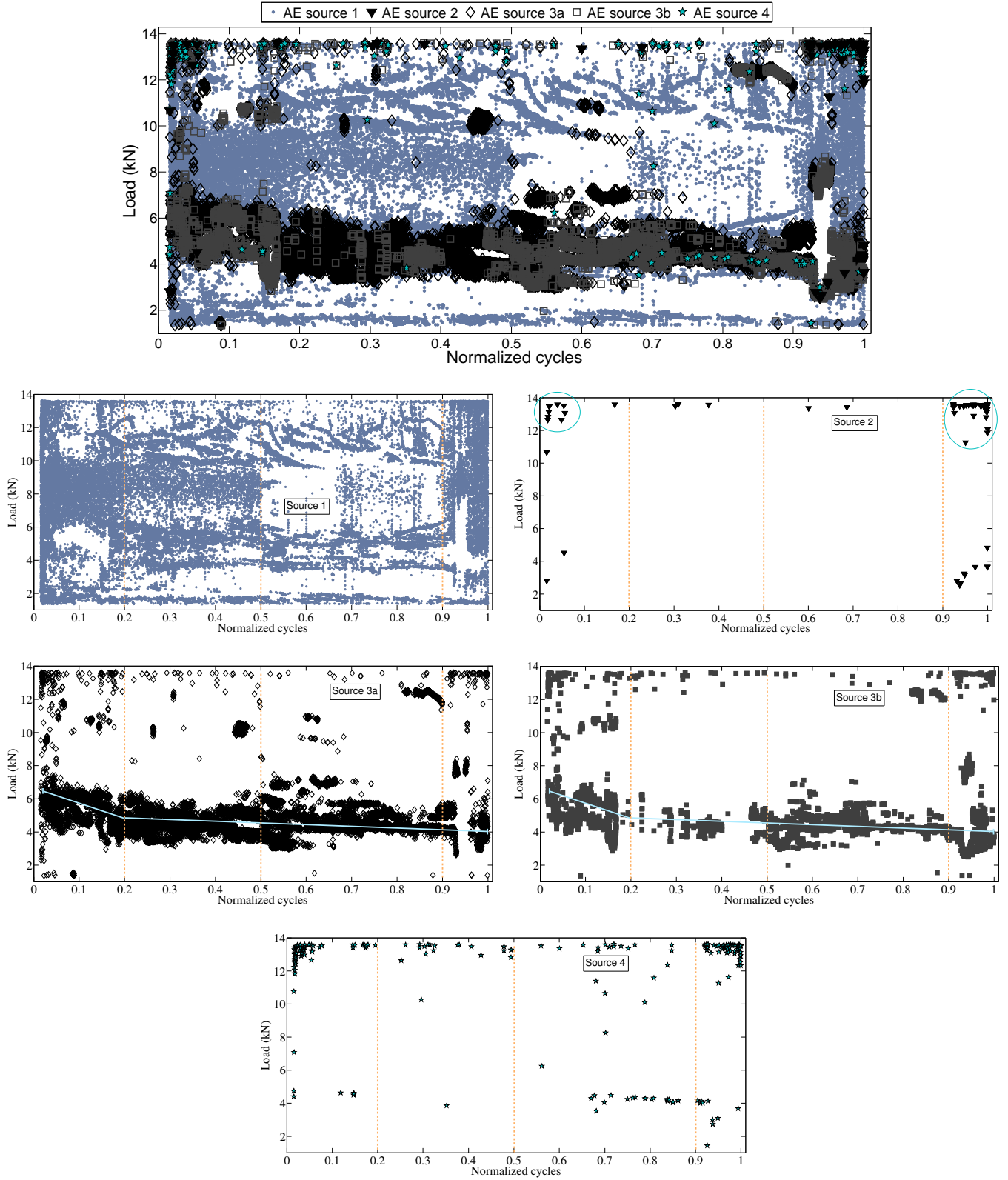
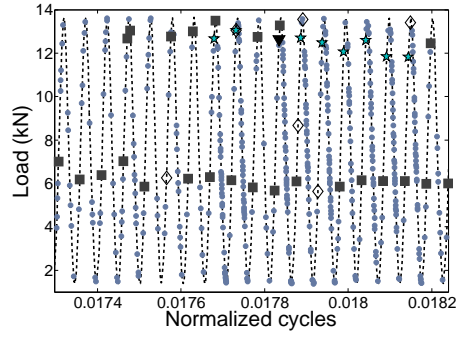
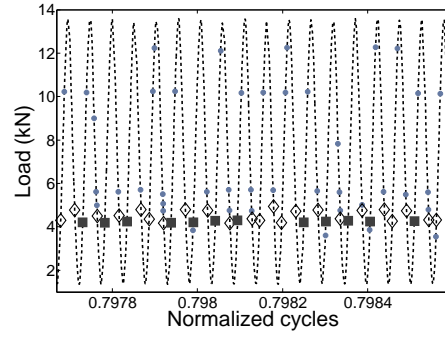


Figure 10: Classified AE events during cyclic loading of specimen A3 (80%): (a) All sources during the whole test; (b)-(f) Individual AE source

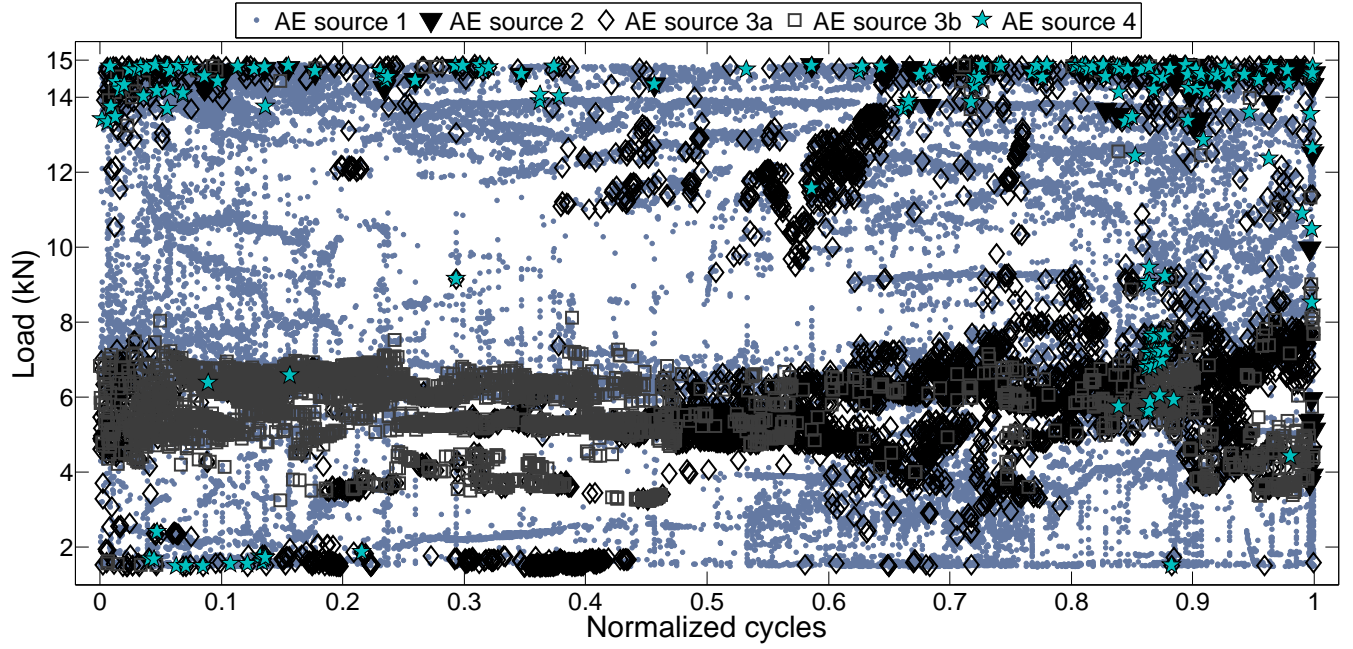


(a)

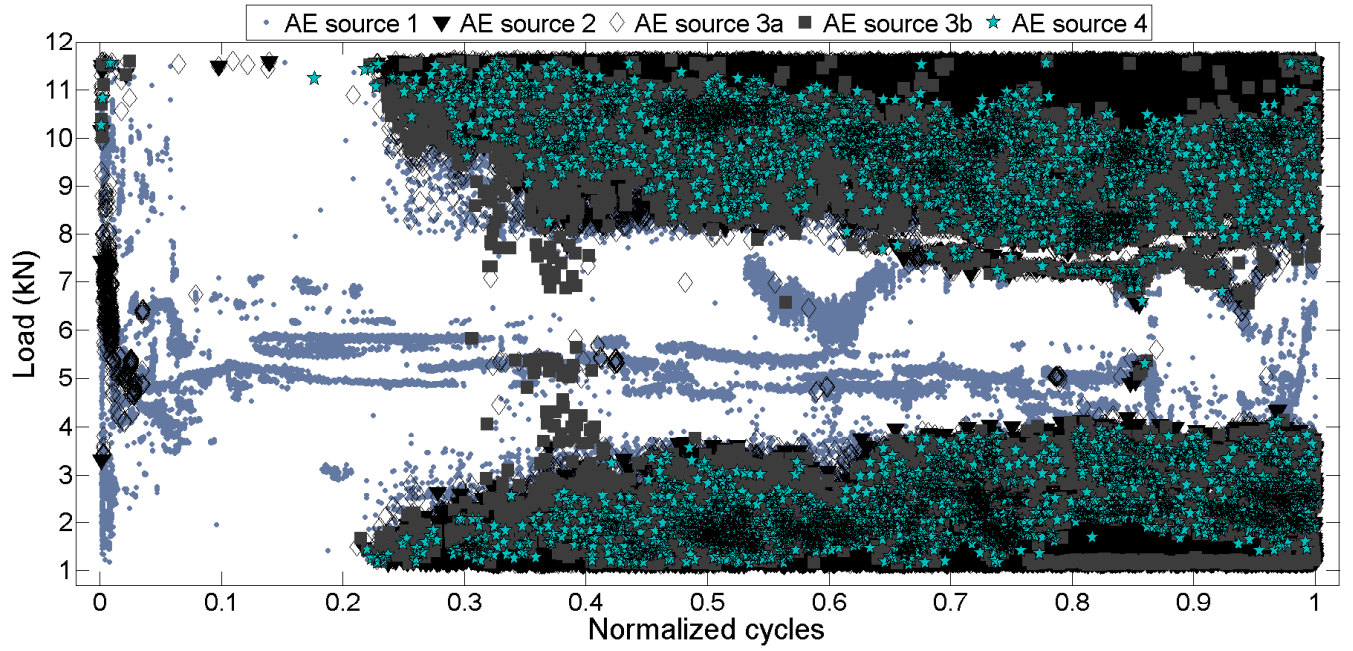


(b)

Figure 11: AE events during cyclic loading of specimen A3 (80%): Close-up view (a) at the beginning and (b) at the end of the test.



(a) Fatigue dataset A2 (90%)



(b) Fatigue dataset A4 (70%)

Figure 12: Visualization of classified AE events during cyclic loading A2 (90% of the ultimate strength) and A4 (70%)

593

List of Tables

594

1

Characteristics of AE datasets considered.

40

595

2

Assigned-to-damage clusters

41

Dataset N/ loading type	% of the ultimate strength	AE hits	Time-to-failure (s)
A1 / quasi-static	x	52,832	0.40E+3
A2 / fatigue	90	481,595	0.74E+3 (3.7E+3 cycles)
A3 / fatigue	80	1,682,434	4.11E+3 (2.0E+4 cycles)
A4 / fatigue	70	9,555,227	2.14E+4 (1.0E+5 cycles)

Table 1: Characteristics of AE datasets considered.

Cluster	AE source
1	Extraneous noise (external friction, hydraulic vibration, EMI)
2	Fiber-related damage (rupture of tows, pull-out)
3a	Friction-related source due to fatigue crack closure under cyclic loading
3b	Matrix-related damage (micro/macro cracking, splitting)
4	Interface-related damage (fiber/matrix)

Table 2: Assigned-to-damage clusters